Foul detection system for race walking events through the integration of multiple sensing functions

Nami Shimizu
Ritsumeikan University
Graduate School of Information Science and Engineering
Osaka, Japan
is0561sp@ed.ritsumei.ac.jp

Hideaki Miyaji
Ritsumeikan University
College of Information Science and Engineering
Osaka, Japan

Hiroshi Yamamoto
Ritsumeikan University
College of Information Science and Engineering
Osaka, Japan

***Abstract*—Race walking is a track and field discipline in which athletes should maintain a state where at least one foot remains in contact with the ground and always keep the supporting leg straight, making its judging criteria stricter than other sports. The judgment is currently performed visually by referees, which requires specialized expertise and often remains difficult for race walking beginners and race walkers who cannot access instructors. In addition, since the population of race walking athletes is relatively small, the number of qualified instructors is limited. Existing studies propose an automated system for detecting foul actions in race walking. For example, pressure sensors are embedded in insoles to detect the time period when both feet leave the ground. However, the reliability of the pressure data deteriorates with the increase in the number of repeated use because the sensors cannot withstand prolonged operation. In addition, AI-based video analysis for pose estimation is proposed to classify the type of foul actions. However, the target area of the method is limited to the camera's field of view and cannot cover the entire field. To address these limitations, this study proposes an automated system for detecting fouls in race walking. In the proposed system, small non-intrusive sensors are attached to the athlete, and the sensor data is analyzed to determine whether the athlete's form satisfies race walking rules. Focusing on the temporal characteristics of the sensor data, the proposed system adopts a Transformer-based model to detect fouls. Furthermore, the system automatically generates training data for the model by linking sensor measurements with accurate posture estimation through AI-based image analysis in limited locations.**

***Index Terms*—Race walking, machine learning, Transformer, pose estimation**

I. Introduction

In recent years, numerous studies are conducted in various sports to analyze data obtained from various sensors and cameras for refereeing and judging [1]. For example, in short-distance running, starting blocks equipped with a sensing system are used to confirm the correctness of the starting conditions. This system measures the pressure applied by athletes to the footplate and calculates the reaction time, which is the time it takes for athletes to respond to the start signal based on pressure changes. In this way, the system can accurately determine the timing of the start and reliably detect false starts [2]. However, in track and field, "race walking" is the only sport event where form is judged exclusively by referees through visual observation. The World Athletics competition rules state that "all the Judges shall act in an individual capacity and their judgments shall be based on observations made by the human eye" [3]. Recently, the organization is advancing the research of sensing technology embedded in insoles to assist with judgment of foul actions in race walking, but the system is under development [4]. For athletes unfamiliar with the competition or for beginners, it is often difficult to judge whether their actions follow the race walking rules.

In the existing studies, various systems are researched and developed to detect fouls by athletes of race walking. Jose P. et al. propose a system that uses pressure sensors to detect the condition where both feet are simultaneously off the ground during race walking [5]. However, the repeated use of the sensors causes the deterioration of the reliability of the sensor data, making them unsuitable for long-term observation. As a non-contact foul detection method, Suzuki et al. propose an AI-based video analysis using a smartphone [6]. However, this method cannot detect foul actions outside the field of view of the camera, limiting the observation range and making it impossible to cover the entire field.

Therefore, in this study, we propose a system that detects foul actions in race walking based on data collected from sensors attached to athletes in a non-intrusive manner. Specifically, the system focuses on the time-series characteristics of sensors and analyzes the sensor data from accelerometer and gyroscope using a Transformer-based machine learning model to detect a typical foul action, loss of contact, which both feet are simultaneously off the ground. In addition, in order to enable the automated construction of training data for machine learning, the system is equipped with depth cameras installed in limited locations on the field. From the captured depth images, the system extracts the joint coordinates of athletes to accurately estimate the occurrence of the foul action. The estimation results are used as reliable ground-truth labels. By linking these labels with sensor data, the system automatically generates training dataset that enables the construction of a robust model. To evaluate the effectiveness of the system, we conduct a proof-of-concept experiment using actual data of race walking athletes.

II. Related Research and Overview of This System

A. Research on sensor-based detection of fouls in race walking

In the existing study by Jose P. et al., a system is proposed that uses pressure sensors to detect the timing when both feet are simultaneously off the ground during race walking [5]. To measure the pressure applied to the soles of the feet, piezoresistive sensors are installed at multiple points on the soles of the shoes. The timing when the pressure (voltage) of all sensors falls below a predetermined threshold is detected as the occurrence of the foul actions, loss of contact. In this system, the pressure sensors produce unreliable data after a certain number of trials, because the sensors can only withstand a limited number of uses.

B. Research on non-wearable approaches to foul detection in race walking

Suzuki et al. propose a method for detecting fouls by analyzing videos recorded by smartphones [6]. By analyzing the videos by utilizing the machine learning model, the positions of joints of the athletes are estimated. And then, the feature values for detecting fouls are calculated from the estimated coordinates of each joint. By receiving the calculated feature vector as an input, the model outputs whether the foul is occurring or not. However, this proposed system has the problem that it cannot detect fouls if the athletes are not within the field of view of the camera.

C. Objective of our study

Existing research has limitations such as the insufficient durability of the sensors and the limitation of observation ranges. Therefore, in this study, we propose a new sensing system that detects loss of contact by analyzing the sensor data from sensor nodes that are attached to the athletes so that their movements are not interfered with.

Specifically, the proposed system analyzes the time-series characteristics of accelerometer and gyroscope data using a Transformer-based machine learning model to detect the foul actions. In addition, to enable the automated construction of the machine learning model, the system utilizes depth cameras installed in limited locations on the field. From the captured depth images, the system extracts joint coordinates of athletes and accurately detects fouls based on the relative positions of both ankles. The detection results based on the depth camera are used as reliable ground-truth labels. By linking these labels with the sensor data, the system automatically generates training data, thereby enabling the automated construction of a robust model capable of detecting foul actions across the entire competition field.

III. Proposed Foul Detection System

A. Overview of the proposed system

Figure 1 shows an overview of the proposed system. As shown in this figure, the proposed system consists of athlete sensor node, form observation node, and an analysis server.

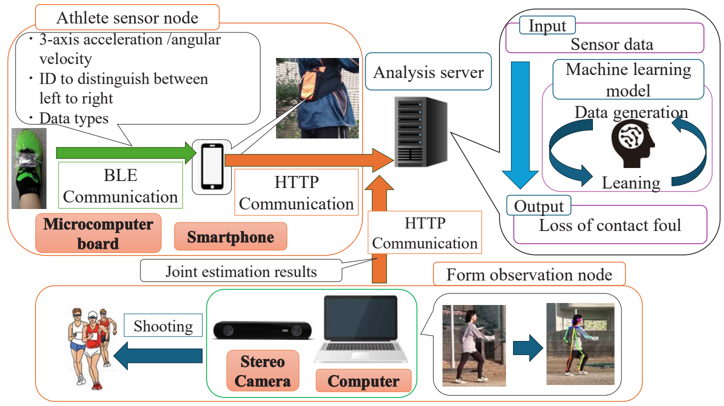


Fig. 1: Overview of proposed system.

The athlete sensor node consists of acceleration sensor and gyroscope sensor worn by the athlete, and a smartphone that transmits the data collected from the sensor to the analysis server. The form observation node consists of a stereo camera that captures depth images of the athletes and a computer that analyzes the images for identifying the athletes' precise form.

The smartphone that is a main component of the athlete sensor node collects sensor data from the microcontroller accommodating the sensors using BLE communication and sends that data to the analysis server. The stereo cameras that make up the form observation node capture the athletes' form and analyze it on the computer connected to the cameras. The computer recognizes the joint coordinates by analyzing the video received from the stereo camera. The analysis server constructs a machine learning model for foul detection by analyzing sensor data received from athlete sensor nodes and the joint coordinate data received from computers, then combining the sensor data with labels indicating the presence or absence of fouls.

B. Hardware configuration of athlete sensor node

The athlete sensor node is worn by athletes. The sensor node consists of a microcontroller board (Feather nRF52840 Sense) which includes various sensors and a communication function by BLE, a lithium-ion battery that powers the microcontroller board, and a smartphone (Redmi Note 10 Pro). To ensure stable communication with the microcontroller board, the smartphone is placed in a waist pouch worn by the athlete. The microcontroller board is placed in an A9-sized zippered plastic bag and is fastened to the shoes with vinyl ties. Figure 2 shows the sensor node worn by the athlete.

C. Hardware configuration of form observation node

The form observation node consists of a stereo camera (ZED 2i) that captures the form of athletes and a notebook PC that performs the joint estimation by analyzing the video captured by the stereo camera. The PC uses the Body Tracking model of the ZED SDK provided by the company that develops the stereo camera to obtain joint coordinate data and sends them to the analysis server via HTTP communication. Based on

Fig. 2: Installation of athlete sensor node.

the time-series of the joint coordinate data, the PC judges whether the foul is occurring or not by analyzing the positional relationship of the joint coordinates of the ankle.

D. Configuration of analysis server

As the analysis server, a standard PC server is utilized. The analysis server analyzes the data transmitted from the athlete sensor node and the form observation node to construct a machine learning model that estimates the correctness of the athletes' form. The constructed model detects foul occurrences and notifies the results. To achieve these functions, Python, scikit-learn, and Tensorflow are installed.

IV. Data analysis for foul detection

A. Sensor data collection flow

The athlete sensor node obtains 3-axis acceleration (m/s^2) and angular velocity (dps (degree per second)) every 20 ms from an acceleration/gyroscope sensor worn by the athlete. The structure of the measured sensor data is shown in Tab. I. In the data structure, the device ID is used to distinguish which foot the sensor is attached to (0: right, 1: left), and the data type ID indicates the type of data being transmitted (acceleration: 0, gyroscope: 65). The smartphone receives the data from the microcontroller through BLE communication and converts that into a JSON-formatted string including a timestamp. And then, the data is sent to the server via HTTP communication. The data structure for the HTTP communication is shown in Tab. II.

TABLE I: Data structure for BLE communication.

Meaning of data	Data size (bytes)	Example
Device ID to distinguish between left and right sensors	2	0
Data type ID representing the type of data	2	0
x-axis sensor data (m/s^2 / dps)	4	9.142523
y-axis sensor data (m/s^2 / dps)	4	2.880120
z-axis sensor data (m/s^2 / dps)	4	5.710553

TABLE II: Data structure for HTTP communication.

Data name	Example	Meaning of data
timestamp	1740802617536	Time when data is received (ms)
deviceID	0	Device ID to distinguish between left and right sensors
dataType	0	Data type ID representing the type of data
x-data	9.921674	x-axis sensor data (m/s^2 / dps)
y-data	2.5314255	y-axis sensor data (m/s^2 / dps)
z-data	5.4414875	z-axis sensor data (m/s^2 / dps)

B. Sensor data analysis for foul detection

The proposed system attempts to detect fouls, loss of contact, in race walking by analyzing time-series data collected from accelerometer and gyroscope sensors attached to the feet. Since walking is essentially a periodic and continuous motion, the temporal characteristics of the sensor data are crucial. The accelerometer data captures the periodic fluctuations corresponding to gait cycles, while the angular velocity data reflects phase changes arising from the rotational movements of the trunk and limbs. By analyzing these continuous variations, early signs of foul actions can be extracted.

To judge the foul actions by analyzing these time-series characteristics effectively, the proposed system adopts a Transformer-based architecture. The Transformer is particularly suitable because of its ability to handle sequential data without relying on strict recurrence, and its attention mechanism enables the model to selectively focus on the most relevant parts of the sequence [7] [8] [9] [10] [11]. This is advantageous for walking analysis, where important parts for the foul detection do not always appear in fixed positions within the data sequence.

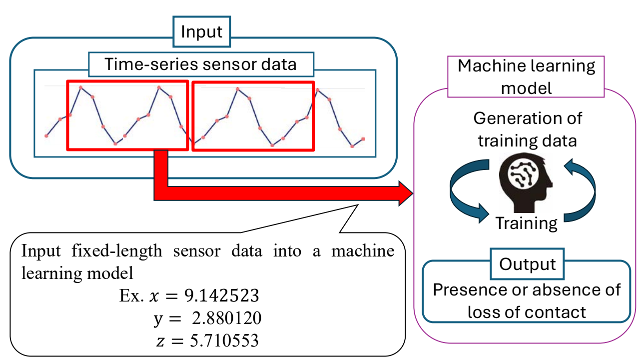


Fig. 3: Overview of machine learning model.

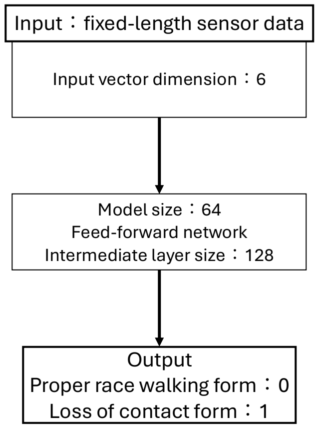


Fig. 4: Machine learning model architecture: Transformer model.

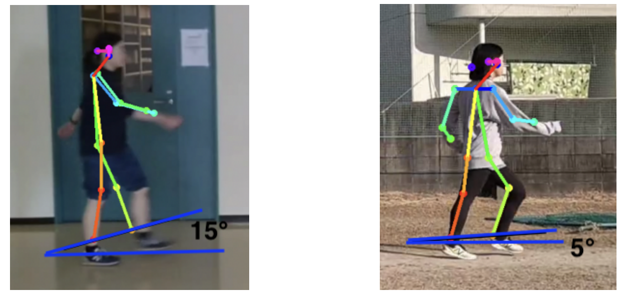
The input to the Transformer model consists of multi-dimensional vectors combining accelerometer and angular velocity features. Specifically, the input vector dimension is six, representing the three-axis accelerometer and three-axis angular velocity data. The Transformer encoder includes one layer, with a model size of 64 and a feed-forward network intermediate layer size of 128. The output layer produces a binary classification label: 0 for "proper race walking form" and 1 for "loss of contact form." Figures 3 and 4 show an overview of the machine learning model.

C. Analysis of depth data for deriving correct ground-truth labels

In race walking, the athlete should walk in such a way that both feet do not leave the ground at the same time. Therefore, the angle of the straight line connecting both ankles is smaller compared to normal walking or running movements. Figures 5, 6 and 7 shows the example of the joint coordinates for each type of movement. From the two examples in Fig. 5, it can be seen that the angle of the line differs depending on the type of movement. Figures 6 and 7 show that the amplitude of the line connecting the right and left ankle joints is large during a loss of contact foul. Accordingly, by using the estimation results of joint coordinates, the inclination of the line connecting the right and left ankle joints is employed as a feature to accurately determine the type of movement.

The analysis server first records the joint coordinate data received from the form observation node into a CSV file. Next, it analyzes each 10 seconds of the time-series data of the angle and counts the number of data points that the angle value falls between -30 degrees and +30 degrees. If the number of data points is 25% or more of the time-series data, the system sets 0 to a label as "proper race walking form for the time-series data." If the number of data points is less than 25%, the system assigns the data with a label of 1 as "loss of contact form."

The data structure for the HTTP communication from the form observation node is shown in Tab. III. The analysis server first receives joint coordinate data and sensor data, classifies them by left, right foot and data type, and saves them in CSV files. These data are then synchronized and integrated based on timestamps, ensuring that sensor data is correctly aligned with joint coordinates. After the integration, the most recent predefined number of synchronized data are grouped as one input data for the machine learning model. By assigning the labels described in Section IV-C to this data, training data is automatically generated. A machine learning model is constructed based on the training data.



(a) Normal Walking (b) Race Walking

Fig. 5: Examples of angle of the line connecting both ankles.

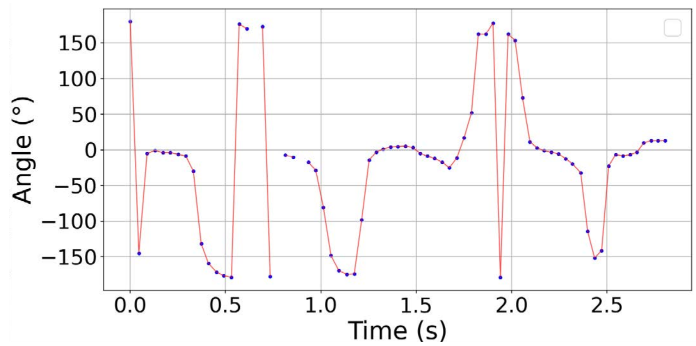


Fig. 6: Time-dependent changes in inclination during race walking motion.

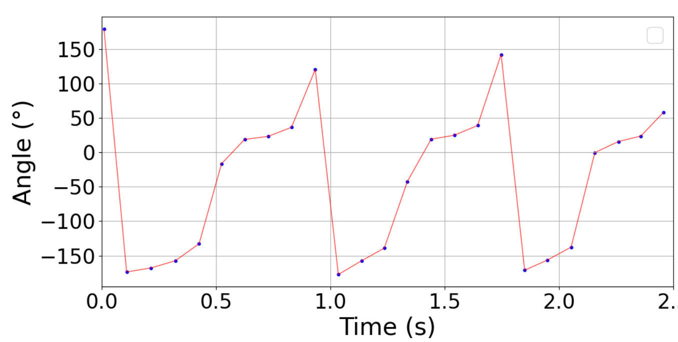


Fig. 7: Time-dependent changes in inclination during loss of contact motion.

TABLE III: List of data and formats sent from stereo camera.

Data name	Example	Meaning of data
id	0	ID for distinguishing individuals
timestamp	1754099456464	Time when data was received (ms)
joint	LEFT_ANKLE	Joint name
dataType	0	Data type ID representing the type of data
x	4.200777053833008	x-coordinate data
y	0.3334699869155884	y-coordinate data
z	4.863706588745117	z-coordinate data

V. PERFORMANCE EVALUATION OF PROPOSED SYSTEM

A. Experimental Setup

In this study, we conduct experiments to evaluate the effectiveness of the proposed foul detection system. The evaluation consists of two parts: (1) a comparison of classification accuracy between manual and automated labeling methods, and (2) a comparison of classification accuracy between the proposed Transformer-based model and an LSTM-based model under different input sequence lengths. In this evaluation, the following performance metrics are utilized:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

In these equations, TP, TN, FP and FN indicate True Positive, True Negative, False Positive and False Negative, respectively.

For data acquisition, a stereo camera mounted on a tripod and a notebook PC are placed together on a movable cart. The cart is moved parallel to the subject to record synchronized video and sensor data. Each recorded video includes a timestamp, and corresponding sensor data are stored in CSV files. The video and sensor data are synchronized based on the timestamps.

For manual labeling, correct labels are assigned through human visual inspection of each frame. For automated labeling, correct labels are assigned using the method described in Section IV-C. A total of 15 sequences of race-walking data and 10 sequences of normal-walking data are used for evaluation.

Four university students (two male and two female) participate in the normal-walking and loss-of-contact trials along a 10 meters indoor corridor. Additionally, three athletes belonging to a university track and field club specializing in race walking participated in the race-walking trials along a 10 meters outdoor course. Each subject performs five trials per motion type while wearing athlete sensor nodes.

B. Evaluation of Labeling Accuracy

To clarify the influence of the labeling method on classification performance, we compared the classification accuracy obtained using manually labeled data with that obtained using automatically labeled data. The results, shown in Fig. 8, indicate that manual labeling achieved 100% accuracy for both correct race-walking and loss-of-contact detection, whereas automated labeling achieved 83%. Although automated labeling achieves slightly lower accuracy compared to manual labeling, the automated labeling achieves high accuracy, demonstrating that the proposed method can effectively generate ground-truth labels without human visual inspection.

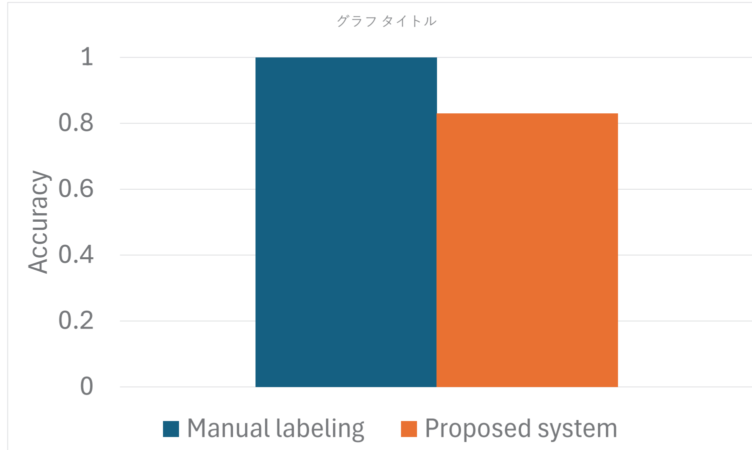


Fig. 8: Accuracy rate of motion estimation.

TABLE IV: Details of training/testing datasets.

	Model 1	Model 2	Model 3
Length of time-series data (ms)	125	250	500
Training data	5400	2700	1360
Test data	4050	2020	1020

C. Evaluation of Model Performance

Next, we evaluate the detection performance of the proposed Transformer-based model by comparing it with an LSTM model, whose architecture is shown in Fig. 9. Both models are trained under multiple sequence-length conditions to examine the effect of input time-series length on detection performance. The number of samples and model configurations are summarized in Tab. IV. As performance metrics, recall, precision, and processing time are utilized.

As shown in Figs. 10 and 11, the increase in the length of the input time-series data improves foul detection accuracy, but also increases processing time. These results clarify that the proposed method is effective for detecting loss-of-contact violations. Moreover, the analysis reveals a trade-off between detection accuracy and computational efficiency, suggesting that the sequence length should be chosen appropriately depending on the application requirements.

D. Evaluation of Processing Efficiency

Figure 12 shows the relationship between processing time and input sequence length. The results indicate that longer input data require greater processing time for inference. Considering both detection accuracy and computational cost, using 250 ms of time-series data as input provides an appropriate balance between accuracy and real-time performance. Overall, these findings demonstrate that the proposed foul detection

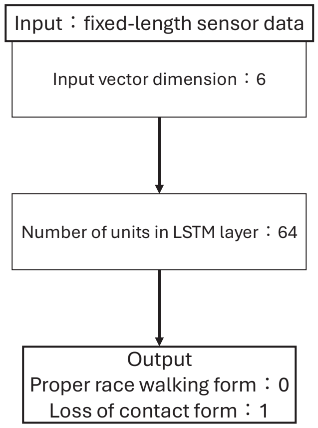


Fig. 9: Machine learning model architecture: LSTM model.

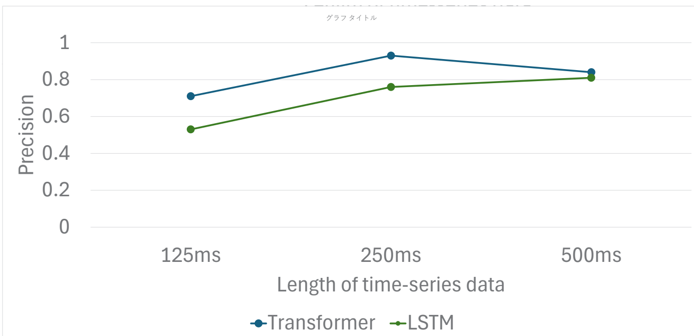


Fig. 10: Precision for classification of foul estimation.

system can achieve high accuracy while maintaining near real-time operation.

VI. Conclusion

In this study, we proposed a sensing system equipped with a machine learning model that detects fouls involving loss of contact in race walking based on observation results from small sensor nodes worn by athletes. In addition, the proposed system has been designed to automatically generate training data for constructing the machine learning model by automatically determining the correct labels for the sensor data. Through the proof-of-concept experiments, it was

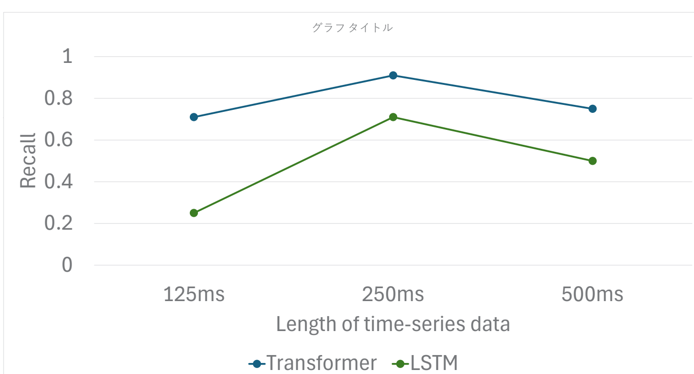


Fig. 11: Recall for classification of foul estimation.

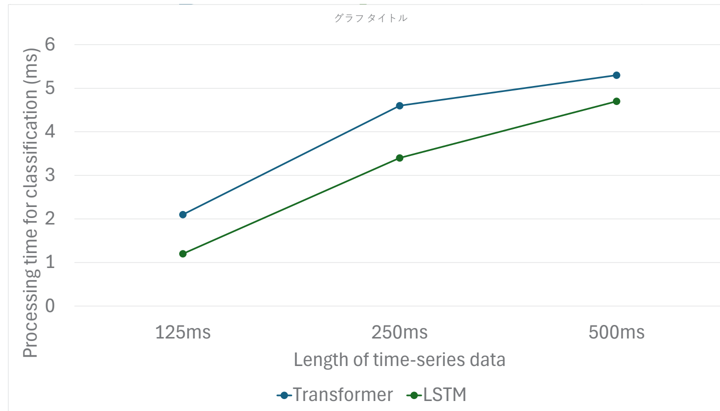


Fig. 12: Processing time for foul estimation.

confirmed that the proposed system can accurately estimate the occurrence of the loss-of-contact fouls. Future challenges include exploring methods to detect the "bent knee" violation in race walking. Additionally, we aim to design a system that simultaneously analyzes the form of multiple athletes and detects foul actions.

Acknowledgment

This work is supported by Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number JP24K20774, JP24K02916, and JP23K28078.

References

[1] B. T. Naik, M. F. Hashmi, and N. D. Bokde, "A Comprehensive Review of Computer Vision in Sports: Open Issues, Future Trends and Research Directions," Applied Sciences, vol. 12, no. 9, pp. 4429-4497, Apr. 2022.

[2] SEIKO TIME CREATION INC., "Start Information System," [Online]. Available: https://www.seiko-sts.co.jp/e/products/detail/sts_792.html. [Accessed: Jul. 24, 2025].

[3] World Athletics, "Competition and Technical Rules 2024 Edition," Jul. 24, 2025.

[4] World Athletics, "Race Walking Committee Makes Recommendations to World Athletics Council," World Athletics, Dec. 1, 2021. [Online]. Available: https://worldathletics.org/news/press-release/race-walking-committee-recommendations?s=06. [Accessed: Sep. 23, 2025].

[5] J. P. Campoverde-Gárate, J. C. Chuqui-Calle, L. J. Serpa-Andrade, and F. L. Bueno-Palomeque, "Detection of Flight Phase in Race Walking Based on Pressure Sensors," in Proc. IEEE 40th Central America and Panama Convention (CONCAPAN), Panama City, Panama, 2022.

[6] T. Suzuki, K. Takeda, and K. Fujii, "Automatic Detection of Faults in Race Walking from a Smartphone Camera: A Comparison of an Olympic Medalist and University Athletes," in Proc. IEEE GCCE, Nov. 2022.

[7] X. Zheng, J. Yin, H. Lu, and J. Chen, "Learning to Estimate Critical Gait Parameters from Single-View RGB Videos with Transformer-Based Attention Network," arXiv preprint arXiv:2312.00398, 2023.

[8] A. Cosma, A. Catruna, and E. Radoi, "Exploring Self-Supervised Vision Transformers for Gait Recognition in the Wild," Sensors, vol. 23, no. 5, pp. 2680-2703, Mar. 2023, doi: 10.3390/s23052680.

[9] C. A. Nakashima, Y. Ma, X. Lin, W. Zhou, and R. Li, "Spatiotemporal Characterization of Gait from Monocular Videos with Transformers," in ICLR 2023 Workshop on Deep Learning for Human Movement Science, Kigali, Rwanda, May 2023.

[10] H. Wang, Y. Hu, and Y. Li, "Gait-ViT: Gait Recognition with Vision Transformer," Sensors, vol. 22, no. 19, pp. 7362-7376, Sep. 2022, doi: 10.3390/s22197362.

[11] Y. Liu, Y. Zhai, and T. Yang, "GaitCycFormer: Leveraging Gait Cycles and Transformers for Gait Emotion Recognition," in Proc. AAAI Conf. Artif. Intell., vol. 38, pp. 2452–2460, 2024.